

# 工作周报总结

2015-06-07

## 本周工作：

1. 智慧城市、农业数据项目（Spark 框架知识储备）
  - Spark 提供的数据集操作类型有很多种，不像 Hadoop 只提供了 Map 和 Reduce 两种操作。比如 map, filter, flatMap, sample, groupByKey, reduceByKey, union, join, cogroup, mapValues, sort, partitionBy 等多种操作类型，他们把这些操作称为 Transformations。同时还提供 Count, collect, reduce, lookup, save 等多种 actions。
  - Spark 的中间数据放到内存中，对于迭代运算效率比较高。
  - 了解 Spark 的 RDD 弹性文件系统的构造流程，了解从 Data Blocks 中构造 RDD 并根据需求调整重新构造 RDD 的步骤。
2. 其余事情
  - 撰写了《Java 代码规范》、《程序开发源代码管理规范》

## 下周工作：

继续熟悉 Spark 框架的知识，对新需求的 AVIS 进行修改。